Compound poisson approximation for dissociated random variables via Stein's method

Eichelsbacher, Peter ; Roos, Malgorzata

Abstract: In the present paper we consider compound Poisson approximation by Stein's method for dissociated random variables. We present some applications to problems in system reliability. In particular, our examples have the structure of an incomplete U-statistics. We mainly apply techniques from Barbour and Utev, who gave new bounds for the solutions of the Stein equation in compound Poisson approximation in two recent papers.

DOI: https://doi.org/10.1017/S0963548399003880

Originally published at:
Eichelsbacher, Peter; Roos, Malgorzata (1999). Compound poisson approximation for dissociated random variables via Stein's method. Combinatorics, Probability Computing, 8(4):335-346.
DOI: https://doi.org/10.1017/S0963548399003880

Compound Poisson Approximation for Dissociated Random Variables via Stein's Method

PETER EICHELSBACHER and MALGORZATA ROOS

Combinatorics, Probability and Computing / Volume 8 / Issue 04 / July 1999, pp 335 - 346
DOI: null, Published online: 08 September 2000

Combinatorics, Probability and Computing (1999) 8, 335–346. Printed in the United Kingdom

Compound Poisson Approximation for Dissociated Random Variables via Stein's Method

PETER EICHELSBACHER[1] and MAŁGORZATA ROOS[2]

[1] Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany
(e-mail: peter@mathematik.uni-bielefeld.de)

[2] Biostatistik, ISPM, Universität Zürich, Sumatrastrasse 30, CH-8006 Zürich, Switzerland
(e-mail: mroos@ifspm.unizh.ch)

Received 5 February 1998; revised 14 September 1998

In the present paper we consider compound Poisson approximation by Stein's method for dissociated random variables. We present some applications to problems in system reliability. In particular, our examples have the structure of an incomplete U-statistics. We mainly apply techniques from Barbour and Utev, who gave new bounds for the solutions of the Stein equation in compound Poisson approximation in two recent papers.

1. Introduction

Let Γ denote an arbitrary finite collection of indices, usually denoted by α, β and so on. Let X_α be 0–1-valued, possibly dependent, random variables and let $W = \sum_{\alpha\in\Gamma} X_\alpha$. If the X_α are weakly dependent and the value 1 occurs with small probability, Poisson approximation in total variation distance between the law of W, denoted by $\mathscr{L}(W)$, and the Poisson distribution with parameter $\mathbb{E}W$, denoted by $\mathrm{Po}(\mathbb{E}W)$, can be successfully established via the Stein–Chen method, which at the same time gives estimates of the approximation error (see Barbour, Holst and Janson [6]). If the random variables X_α can take other positive integer values or if the dependence is stronger (clumps of 1s tend to occur), the compound Poisson distribution should provide better approximation. The importance of developing a compound Poisson approach has been discussed in Aldous [1].

In [3], Barbour, Chen and Loh introduced a Stein equation for compound Poisson approximation. Roos developed the local version of the basic method in [12] and in [11] the coupling approach. The theoretical results were successfully applied to many examples in reliability theory (see [4]). The solutions of the corresponding Stein equation may in general grow exponentially with the mean number of clumps. For a special class of distributions in [3, Theorem 5] a better bound was given which is comparable in sharpness to the corresponding bound in the Poisson Stein–Chen method (apart from a logarithmic term in the numerator of the bound). Recently, Barbour and Utev [8] proved bounds for the solutions of the Stein equation with respect to the Kolmogorov distance which enabled them to carry out compound Poisson approximation in this distance with the same efficiency as in the Poisson case. They assumed that the compound Poisson limit distribution has a fourth moment and that it satisfies an aperiodicity condition. In [7] the same authors proved the counterparts in total variation distance under the assumption that the compound Poisson limit distribution is aperiodic of finite exponential moment, which is hardly restrictive. The contribution of [7] and [8] is fundamental.

The aim of the present paper is to improve some results stated in [4, 12] by using the new bounds from [7] and [8]. We observe that the random variables in the examples considered up to now are actually *dissociated* ones. Moreover, we present Stein's method for compound Poisson approximation both in Kolmogorov distance and in total variation distance for this special class of random variables. In view of the theoretical work by Barbour and Utev we provide the following examples and applications: k-runs, isolated vertices in the rectangular lattice on the torus and the two-dimensional consecutive-k-out-of-n system. In particular all the examples have the structure of an incomplete U-statistics. After having calculated the bound for compound Poisson approximation for U-statistics, we see that improvements of the Poisson approximation results on U-statistics (see [5]) cannot be expected in general. In the examples we are able to obtain estimates with the right behaviour when the mean number of clumps tends to infinity. As a consequence of the good bounds in [7] and [8] no unwanted logarithmic factor and no unwanted e^{λ}, where λ is the mean number of clumps, come into play. In Section 2, we state known theoretical results about Stein's method and give the theorem for dissociated random variables as well as *locally dependent* ones. In Section 3 examples are presented.

2. The bounds

2.1. Stein's method

We denote by $CP(\lambda, \boldsymbol{\mu})$ the compound Poisson distribution of $\sum_{j=1}^{N} X_j$, where $(X_j, j \geqslant 1)$ are nonnegative integer-valued, independent and identically distributed variables with distribution $\boldsymbol{\mu}$ (a probability measure on $(0,\infty)$) and are independent of N, distributed according to the usual Poisson distribution with mean λ. We briefly summarize Stein's method for this class of distributions. If W is a nonnegative integer-valued random variable satisfying

$$\left|\mathbb{E}\left(\sum_{i\geqslant 1} i\lambda_i g(W+i) - Wg(W)\right)\right| \leqslant \varepsilon_0\, M_0(g) + \varepsilon_1\, M_1(g) \tag{2.1}$$

for all bounded $g : \mathbb{N} \to \mathbb{R}$ and for some small ε_0 and ε_1, where

$$M_0(g) = \sup_{j\geqslant 1} |g(j)|, \qquad M_1(g) = \sup_{j\geqslant 1} |g(j+1) - g(j)|,$$

then it follows that

$$d_{TV}(\mathscr{L}(W), CP(\lambda, \boldsymbol{\mu})) \leqslant \varepsilon_0\, H_0(\lambda, \boldsymbol{\mu}) + \varepsilon_1\, H_1(\lambda, \boldsymbol{\mu}), \tag{2.2}$$

where $\lambda = \sum_{i\geqslant 1} \lambda_i$, $\mu_i = \lambda_i/\lambda$ (that means that the measure $\boldsymbol{\mu}$ has the form $\boldsymbol{\mu} = \sum_{i\geqslant 1} \mu_i\,\delta_i$, where δ_i denotes the Dirac measure concentrated on $i \in \mathbb{N}$), $d_{TV}(P,Q)$ denotes the total variation distance between two probability measures P and Q on $\mathbb{Z}_+$ ($d_{TV}(P,Q) = \sup_{A\subseteq\mathbb{Z}_+}\{|P(A)-Q(A)|\}$) and $H_l(\lambda,\boldsymbol{\mu}) = \sup_{A\subset\mathbb{Z}_+} M_l(g_A)$, $l = 0,1$, with g_A being the solution of the Stein equation

$$\sum_{i\geqslant 1} i\lambda_i g(w+i) - wg(w) = 1_A(w) - CP(\lambda,\boldsymbol{\mu})\{A\}, \quad w \geqslant 0,\ A \subseteq \mathbb{Z}_+. \tag{2.3}$$

The method works as follows: if we can bound the left-hand side of (2.1) with suitably small ε_0 and ε_1 and the constants $H_l(\lambda,\boldsymbol{\mu})$ are not too large, the distribution of W is close to an appropriate compound Poisson distribution. The bounds on $H_l(\lambda,\boldsymbol{\mu})$ will depend on λ and $\boldsymbol{\mu}$ and not on the given random variable W itself.

There may not be any good bounds for the $H_l(\lambda,\boldsymbol{\mu})$ (see discussion in the introduction of [7]), but for the terms

$$H_l^{(a)}(\lambda,\boldsymbol{\mu}) = \sup_{A\subset\mathbb{Z}_+} M_l(g_A(\cdot + a)),\ l = 0,1,$$

and, by inequality (4.4) in [8], (2.2) can be replaced by

$$\begin{aligned} d_{TV}(\mathscr{L}(W), CP(\lambda,\boldsymbol{\mu})) \leqslant{}& \varepsilon_0 H_0^{(a)}(\lambda,\boldsymbol{\mu}) + \varepsilon_1\left(H_1^{(a)}(\lambda,\boldsymbol{\mu}) + \frac{2H_0^{(a)}(\lambda,\boldsymbol{\mu})}{\lambda m_1(1-c_1)}\right) \\ &+\mathbb{P}\left(W \leqslant \frac{1}{2}(1+c_1)\lambda m_1\right)\left(1 + \frac{2m_2 H_0^{(a)}(\lambda,\boldsymbol{\mu})}{m_1(1-c_1)}\right), \end{aligned} \tag{2.4}$$

where $c_1 \in (0,1)$, $a = c_1\lambda m_1$ and m_l denotes $\sum_{i\geqslant 1} i^l\,\mu_i$ for $l \geqslant 1$. Let us assume that the following two conditions hold.

Condition 2.1. $\boldsymbol{\mu}(z) = \sum_{i\geqslant 1} \mu_i\, z^i$ *has radius of convergence* $R(\boldsymbol{\mu}) > 1$ (*meaning that we assume that* $\boldsymbol{\mu}$ *has a finite exponential moment*), $\lambda \geqslant 2$.

Condition 2.2 (Aperiodicity). *For all* $0 < \zeta \leqslant \pi : \varrho^*(\zeta) > 0$, *where* $\varrho^*(\zeta) = \min(\varrho_1^*(\zeta), \frac{1}{2}\varrho_2^*(\zeta), 1)$ *with* $\varrho_k^*(\zeta) = \inf_{\zeta\leqslant\theta\leqslant\pi} \varrho_k(\theta)$, $k = 1,2$, *and* $\varrho_1(\theta) = 1 - \sum_{i\geqslant 1} \mu_i \cos i\theta$ *and* $\varrho_2(\theta) = 1 - \frac{1}{m_1}\sum_{i\geqslant 1} i\mu_i \cos i\theta$ (*note that the condition implies* $\boldsymbol{\mu}\{l\,\mathbb{Z}_+\} < 1$ *for any* $l \geqslant 2$).

Then there exist constants $C_l(\boldsymbol{\mu})$, $l = 0,1,2$, given explicitly in terms of $\boldsymbol{\mu}$ (see [7, (1.20)–(1.28)]) such that, for any $a \geqslant C_2(\boldsymbol{\mu})\lambda m_1 + 1$,

$$H_0^{(a)}(\lambda,\boldsymbol{\mu}) \leqslant \frac{1}{\sqrt{\lambda}} C_0(\boldsymbol{\mu}), \qquad H_1^{(a)}(\lambda,\boldsymbol{\mu}) \leqslant \frac{1}{\lambda} C_1(\boldsymbol{\mu}). \tag{2.5}$$

Note that $C_2(\boldsymbol{\mu}) = 1 - \varrho^*(\zeta_0)/4$ with $\zeta_0 = \sqrt{m_2/m_4}$. Because $C_2(\boldsymbol{\mu}) < 1$ there are feasible choices for c_1 in (2.4). In [8], where the Kolmogorov distance is considered, the $H_l(\lambda,\boldsymbol{\mu})$ are replaced by

$$J_l(\lambda,\boldsymbol{\mu}) = \sup_{m\geqslant 0} M_l(g_{[0,m]}),\ l = 0,1,$$

which play the same role as the $H_l(\lambda,\boldsymbol{\mu})$ if one considers the smaller family of sets

$A = [0, m]$, $m \geqslant 1$, on the right-hand side of the Stein equation (2.3). Then the estimate for approximation in terms of Kolmogorov distance is the analogue of (2.2):

$$d_K\big(\mathscr{L}(W), CP(\lambda, \boldsymbol{\mu})\big) \leqslant \varepsilon_0 \, J_0(\lambda, \boldsymbol{\mu}) + \varepsilon_1 \, J_1(\lambda, \boldsymbol{\mu}),$$

where $d_K(\mathscr{L}(W), CP(\lambda, \boldsymbol{\mu})) = \sup_{t \in \mathbb{R}} |\mathbb{P}(W \leqslant t) - CP(\lambda, \boldsymbol{\mu})\{(-\infty, t)\}|$. The quantities $J_l^{(a)}(\lambda, \boldsymbol{\mu})$, $l = 0, 1$, are obtained in the same way as the $H_l^{(a)}(\lambda, \boldsymbol{\mu})$ but with $g_{[0,m]}$ instead of g_A, $A \subseteq \mathbb{Z}_+$. Under weaker assumptions, namely Condition 2.2 and the finiteness of the fourth moment, Theorem 4.3 in [8] gives a bound in the Kolmogorov distance:

$$\begin{aligned} d_K(\mathscr{L}(W), CP(\lambda, \boldsymbol{\mu})) \leqslant \; & \varepsilon_0 \, J_0^{(a)}(\lambda, \boldsymbol{\mu}) + \varepsilon_1 \left(J_1^{(a)}(\lambda, \boldsymbol{\mu}) + \frac{4 J_0^{(a)}(\lambda, \boldsymbol{\mu})}{\lambda m_1 \varrho^*(\zeta_0)} \right) \\ & + \mathbb{P}\big(W \leqslant \lambda m_1 (1 - \varrho^*(\zeta_0)/4)\big) \left(1 + \frac{4 m_2 J_0^{(a)}(\lambda, \boldsymbol{\mu})}{m_1 \varrho^*(\zeta_0)} \right), \end{aligned} \tag{2.6}$$

where $\zeta_0 = \sqrt{m_2/m_4}$, ϱ^* is as defined above and the bounds for $J_0^{(a)}(\lambda, \boldsymbol{\mu})$ and $J_1^{(a)}(\lambda, \boldsymbol{\mu})$ are stated later.

2.2. Dependence structure

Suppose that Γ is a collection of k-subsets $\alpha = \{\alpha_1, \dots, \alpha_k\}$ of $\{1, 2, \dots, n\}$. Then a family $(X_\alpha, \alpha \in \Gamma)$ of nonnegative integer-valued random variables is said to be *dissociated* if the subsets of random variables $(X_\alpha, \alpha \in A)$ and $(X_\beta, \beta \in B)$ are independent whenever $(\cup_{\alpha \in A} \alpha) \cap (\cup_{\beta \in B} \beta) = \emptyset$. Now we state a result on bounds for ε_0 and ε_1 in the case of nonnegative integer-valued dissociated random variables $(X_\alpha, \alpha \in \Gamma)$. We require some further notation. For Γ fixed, we define the *dissociated partition* for each $\alpha \in \Gamma$:

$$\Gamma = \{\alpha\} \cup \Gamma_\alpha^{vs} \cup \Gamma_\alpha^0 \cup \Gamma_\alpha^w$$

and set

$$Z_\alpha = \sum_{\beta \in \Gamma_\alpha^{vs}} X_\beta, \qquad U_\alpha = \sum_{\beta \in \Gamma_\alpha^0} X_\beta, \qquad W_\alpha = \sum_{\beta \in \Gamma_\alpha^w} X_\beta.$$

Here Γ_α^{vs} contains indices of those X_β which strongly influence X_α. Depending on the choice of this region we define for each α

$$\Gamma_\alpha^0 = \{\beta \in \Gamma : \beta \notin \Gamma_\alpha^{vs} \cup \{\alpha\}, \beta \cap \delta \neq \emptyset \quad \text{for some} \quad \delta \in \Gamma_\alpha^{vs} \cup \{\alpha\}\}$$

and

$$\Gamma_\alpha^w = \{\beta \in \Gamma : \beta \cap \delta = \emptyset \quad \text{for all} \quad \delta \in \Gamma_\alpha^{vs} \cup \{\alpha\}\}.$$

Note that the random variables with indices in the region Γ_α^w do not influence X_α and Z_α. Moreover, the definition of Γ_α^{vs} provides us with enough freedom of choice in defining the clumps. In particular, the random variables with indices in Γ_α^0 may be dependent on X_α.

Theorem 2.3. *If $(X_\alpha, \alpha \in \Gamma)$ are nonnegative integer-valued dissociated random variables, Γ_α^{vs}, Γ_α^0 and Γ_α^w are as defined above, then we can take $\varepsilon_0 = 0$ and*

$$\varepsilon_1 = \sum_{\alpha \in \Gamma} \Big(\mathbb{E}(X_\alpha) \, \mathbb{E}(X_\alpha + Z_\alpha + U_\alpha) + \mathbb{E}(X_\alpha \, U_\alpha) \Big). \tag{2.7}$$

Proof. We only have to choose the special dissociated partition $\Gamma_\alpha^b = \Gamma_\alpha^0$ in Lemma 1.8 in [7]. Since the X_α are dissociated, W_α is independent of Z_α and X_α, and, therefore, $\delta_1 = \delta_2 = \delta_3 = 0$ and $\delta_4 = \varepsilon_1$. The theorem follows immediately. □

Remark 1. There are cases when one is able to find some probabilistic structure in Γ_α^0 in the case of dissociated random variables. In such situation it might be advantageous to use the coupling approach instead of the local one: see Remark 2.1.5 in [6].

A family of nonnegative integer-valued random variables $(X_\alpha, \alpha \in \Gamma)$ is said to be *locally dependent* if, for each $\alpha \in \Gamma$, there exist $A_\alpha \subseteq B_\alpha \subseteq \Gamma$ with $\alpha \in A_\alpha$ such that X_α is independent of $(X_\beta, \beta \in A_\alpha^c)$ and $(X_\beta, \beta \in A_\alpha)$ is independent of $(X_\beta, \beta \in B_\alpha^c)$. It is well known (see [3, Section 4]) that m-dependence and finite dependence are special cases of local dependence. If Γ is a collection of k-subsets then a locally dependent family $(X_\alpha, \alpha \in \Gamma)$ is also a dissociated family by definition: for a fixed α we choose $A_\alpha = \{\beta \in \Gamma : \alpha \cap \beta \neq \emptyset\}$ and $B_\alpha = A_\alpha \cup \{\beta \in \Gamma : \beta \notin A_\alpha \quad \text{and} \quad \beta \cap \delta \neq \emptyset \text{ for some } \delta \in A_\alpha\}$. Note that, by definition, $\alpha \in A_\alpha$. Actually a dissociated family is also a locally dependent family. For a locally dependent family we obtain the following.

Corollary 2.4. *If* $(X_\alpha, \alpha \in \Gamma)$ *are nonnegative integer-valued locally dependent random variables, then we can take* $\varepsilon_0 = 0$ *and*

$$\varepsilon_1 = \sum_{\alpha \in \Gamma} \sum_{\beta \in B_\alpha} \mathbb{E}X_\alpha \, \mathbb{E}X_\beta + \sum_{\alpha \in \Gamma} \sum_{\beta \in B_\alpha \setminus A_\alpha} \mathbb{E}X_\alpha \, \mathbb{E}X_\beta. \tag{2.8}$$

Proof. For any $\alpha \in \Gamma$ we choose the partition $\{\alpha\} \cup \Gamma_\alpha^{vs} = A_\alpha$, $\Gamma_\alpha^b = B_\alpha \setminus A_\alpha$ and $\Gamma_\alpha^w = B_\alpha^c$ in Theorem 2.3. By the definition of local dependence X_α and X_β with $\beta \in B_\alpha \setminus A_\alpha$ are independent. □

Remark 2. Corollary 2.4 is an improvement of Theorem 8 in [3]: we are dealing with the case of nonnegative integer-valued random variables. In [3] a slightly weaker bound $\varepsilon_1 = 2 \sum_{\alpha \in \Gamma} \sum_{\beta \in B_\alpha} \mathbb{E}X_\alpha \, \mathbb{E}X_\beta$ is proved.

Remark 3. Note that the approximating distribution $CP(\lambda, \boldsymbol{\mu})$ can be calculated explicitly: $\mu_i = \frac{1}{i\lambda} \sum_{\alpha \in \Gamma} \mathbb{E}(X_\alpha I[X_\alpha + Z_\alpha = i])$, $i \geqslant 1$, $\lambda = \sum_{\alpha \in \Gamma} \mathbb{E}\big(\frac{X_\alpha}{X_\alpha + Z_\alpha} I[X_\alpha + Z_\alpha \geqslant 1]\big)$.

One can consider an approximation by a $CP(\lambda, \boldsymbol{\mu}')$ distribution instead of the naturally emerging distribution $CP(\lambda, \boldsymbol{\mu})$. Theorem 1.10 in [7] states that for any choice of $\boldsymbol{\mu}'$ one gets (2.4), where $\boldsymbol{\mu}$, m_1, m_2 are replaced by $\boldsymbol{\mu}'$, m_1', m_2' and ε_0 and ε_1 are replaced by $\varepsilon_0' = \lambda |m_1 - m_1'| + \varepsilon_0$ and $\varepsilon_1' = \lambda m_1 \, d_W(\boldsymbol{\mu}^\star, \boldsymbol{\mu}'^\star) + \varepsilon_1$; here * is used to denote a size-biased distribution $\mu_i^\star = i\mu_i / m_1$ and d_W denotes the Wasserstein L_1 metric on probability measures over $\mathbb{R}$. Note that under dissociation and local dependence, when $\varepsilon_0 = 0$, the approximation by $CP(\lambda, \boldsymbol{\mu}')$ gives us a term $\lambda |m_1 - m_1'| H_0^{(a)}(\lambda, \boldsymbol{\mu}')$ which may have the biggest contribution to the order of the upper bound. Instead, one might think about another approximation distribution $CP(\lambda^*, \boldsymbol{\mu}^*)$ (which modifies both distribution $\boldsymbol{\mu}$ and the mean number of clumps λ). In doing so it is easier to deal with parameters $\lambda_i^* = \lambda^* \mu_i^*$. We will follow suggestions and proofs of Theorem 3.B in [10] and Theorem 3 in [12] together with the proof of Lemma 1.8 and Theorem 1.10 in [7]. For any choice of the index set Γ_α^{vs}, define $D := \max_{\alpha\in\Gamma}\{|\Gamma_\alpha^{vs}|\}$.

Theorem 2.5. *Let $\lambda_i \geqslant 0$ for $i \in \{1,2,\dots,D+1\}$. For a fixed number $l < D+1$, let $\lambda_1^* = \lambda_1 + \sum_{i=l+1}^{D+1} i\,\lambda_i$, $\lambda_i^* = \lambda_i$ for $i = 2,\dots,l$ and $\lambda_i^* = 0$ for $i \geqslant l+1$. Then*

$$d_{TV}(\mathscr{L}(W), CP(\lambda^*,\boldsymbol{\mu}^*)) \leqslant \varepsilon_0\, H_0^{(a)}(\lambda^*,\boldsymbol{\mu}^*) + \varepsilon_1^*\left(H_1^{(a)}(\lambda^*,\boldsymbol{\mu}^*) + \frac{2H_0^{(a)}(\lambda^*,\boldsymbol{\mu}^*)}{\lambda^* m_1^*(1-c_1)}\right)$$
$$+\mathbb{P}\left(W \leqslant \frac{1}{2}(1+c_1)\lambda^* m_1^*\right)\left(1 + \frac{2m_2^* H_0^{(a)}(\lambda^*,\boldsymbol{\mu}^*)}{m_1^*(1-c_1)}\right), \quad (2.9)$$

where ε_0 stays unchanged and $\varepsilon_1^* = \varepsilon_1 + \sum_{i=l+1}^{D+1} i(i-1)\lambda_i$.

Proof. Let $\tilde{\lambda}_1 = \lambda_1$ and $\lambda_i - \tilde{\lambda}_i \geqslant 0$ for $i \geqslant 2$ and let g be the solution of the Stein equation (2.3) of the function $1_A(w) - CP(\lambda^*,\boldsymbol{\mu}^*)\{A\}$. We obtain

$$\begin{aligned}\sum_{i=1}^{D+1} i\lambda_i g(W+i) - Wg(W) &= \sum_{i=1}^{D+1} i\tilde{\lambda}_i g(W+i) - Wg(W)\\ &\quad+\sum_{i=1}^{D+1} i(\lambda_i-\tilde{\lambda}_i)(g(W+i)-g(W+1)) + \sum_{i=1}^{D+1} i(\lambda_i-\tilde{\lambda}_i)g(W+1)\\ &= \left(\tilde{\lambda}_1 + \sum_{i=1}^{D+1} i(\lambda_i-\tilde{\lambda}_i)\right)g(W+1) + \sum_{i=2}^{D+1} i\tilde{\lambda}_i g(W+i) - Wg(W)\\ &\quad+\sum_{i=2}^{D+1} i(\lambda_i-\tilde{\lambda}_i)(g(W+i)-g(W+1)).\end{aligned} \quad (2.10)$$

Now define $\tilde{\lambda}_1 = \lambda_1$ and $\tilde{\lambda}_i = \lambda_i$ for $i = 2,\dots,l$ and $\tilde{\lambda}_i = 0$ for $i \geqslant l+1$ and denote $\lambda_1^* = \tilde{\lambda}_1 + \sum_{i=1}^{D+1} i(\lambda_i - \tilde{\lambda}_i) = \lambda_1 + \sum_{i=l+1}^{D+1} i\lambda_i$ and $\lambda_i^* = \tilde{\lambda}_i$ for $i = 2,\dots,D+1$. Thus from (2.10) we obtain

$$\left|\mathbb{E}\left\{\sum_{i=1}^{D+1} i\,\lambda_i^*\, g(W+i) - W\,g(W)\right\}\right| \leqslant$$
$$\left|\mathbb{E}\left\{\sum_{i=1}^{D+1} i\,\lambda_i\, g(W+i) - W\,g(W)\right\}\right| + \sum_{i=l+1}^{D+1} i(i-1)\lambda_i\, M_1(g).$$

Combining this with (2.4) we get the result. □

Remark 4. Note that $\boldsymbol{\mu}^*$ has a smaller set of possible values $\{1,\dots,l\}$ ($l < D+1$) instead of $\{1,\dots,D+1\}$ and $\lambda^* = \sum_{i\geqslant 1}\lambda_i^* \neq \sum_{i=1}^{D+1}\lambda_i = \lambda$.

Remark 5. There are different approaches to bounding the truncation term $\mathbb{P}\big(W \leqslant (1-\varepsilon)\mathbb{E}W\big)$: for example, one might use the Chebyshev inequality. In our examples we mostly apply Janson's inequality, which works for indicator random variables built up from independent ones. We state the inequality (see [6, Theorem 2.S]). Consider a collection of independent random indicator variables $(J_i, i \in Q)$ and a finite family of subsets $(Q(\alpha), \alpha \in \Gamma)$ of the index set Q and define $I_\alpha = \prod_{i \in Q(\alpha)} J_i$ and $W = \sum_{\alpha \in \Gamma} I_\alpha$. Partition Γ into $\Gamma_\alpha^+ \cup \Gamma_\alpha^i$, where $\Gamma_\alpha^+ = \{\beta \neq \alpha : Q(\alpha) \cap Q(\beta) \neq \emptyset\}$, and define $\delta = \frac{1}{\mathbb{E}W} \sum_{\alpha \in \Gamma} \sum_{\beta \in \Gamma_\alpha^+} \mathbb{E} I_\alpha I_\beta$. Then, for any $0 \leqslant \varepsilon \leqslant 1$, Janson obtained

$$\mathbb{P}\big(W \leqslant (1-\varepsilon)\mathbb{E}W\big) \leqslant \exp\left(-\frac{1}{2(1+\delta)}\varepsilon^2\, \mathbb{E}W\right). \tag{2.11}$$

3. Applications

3.1. *k*-runs

Consider the problem of k-runs of 1s in a series of independent identically distributed Bernoulli random variables. In [2, Section 4.2.1], when approximating the distribution of the number of overlapping k-runs of heads, Arratia, Goldstein and Gordon give a bound on the total variation distance from a compound Poisson approximation of order $nkp^{2k}(1-p)$ and Roos improves this in [10] to order $kp^k \log(np^k)$ when $p < 1/2$. Suppose n satisfies $n > 4k-3$. Consider independent random variables $(J_j, j \in \{1, \ldots, n\})$ with $\mathbb{P}(J_j = 1) = p = 1 - \mathbb{P}(J_j = 0)$, $0 < p < 1$. To avoid edge effects assume that indices of the form $i + nl$ are identified with i whenever $1 \leqslant i \leqslant n$ and $l \in \mathbb{Z}$. Define the family $(Q(\alpha), \alpha \in \Gamma)$ of subsets of index set $\{1, \ldots, n\}$ consisting of k consecutive indices. We identify each index set α consisting of k consecutive indices with the first index on its left-hand side. Thus $|\Gamma| = n$. We define $I_\alpha = \prod_{j=\alpha}^{\alpha+k-1} J_j$ and $W = \sum_{\alpha \in \Gamma} I_\alpha$. Then $\mathbb{E}I_\alpha = p^k$ and $\mathbb{E}W = np^k$. The random variables $(I_\alpha, \alpha \in \Gamma)$ are dissociated. Choose $\Gamma_\alpha^{vs} = \{\alpha-(k-1), \ldots, \alpha-1, \alpha+1, \ldots, \alpha+k-1\}$ and Γ_α^0 and Γ_α^w as in the dissociated partition. Then, applying Theorem 2.3, we obtain $\varepsilon_1 = (6k-5)n\,p^{2k}$. Furthermore, we will calculate the limiting measure $CP(\lambda, \boldsymbol{\mu})$; see Remark 3. As in [10, Lemma 3.3.4] we obtain that

$$\lambda_i = \begin{cases} np^k p^{i-1}(1-p)^2, & \text{for } i = 1, \ldots, k-1, \\ \frac{np^k}{i}(2p^{i-1}(1-p) + (2k-i-2)p^{i-1}(1-p)^2), & \text{for } i = k, \ldots, 2k-2, \\ \frac{np^k}{2k-1} p^{2k-2}, & \text{for } i = 2k-1. \end{cases}$$

Thus we can calculate $\lambda = \sum_{i=1}^{2k-1} \lambda_i$, the μ_is, and the moments of $\boldsymbol{\mu}$. For the discussions we will use the fact that $\lambda \asymp n\,p^k\,(1-p+p^k)$. It is obvious that m_4 is finite and Condition 2.2 is fulfilled, since $\mu_1 = \lambda_1/\lambda = np^k(1-p)^2/\lambda \geqslant (1-p)^2 > 0$ (see [8, Remark before Lemma 4.1]). So we can apply (2.6): therefore, we have to determine the order of $J_l^{(a)}(\lambda, \boldsymbol{\mu})$, $l = 0, 1$, and to calculate the remaining contribution $\mathbb{P}\big(W \leqslant \lambda m_1(1 - \varrho^*(\zeta_0)/4)\big)$ using an appropriate inequality. As stated in [8, Theorem 4.3], choosing $a = (1 - \varrho^*(\zeta_0)/2)\lambda m_1$ one obtains that $J_0^{(a)}(\lambda, \boldsymbol{\mu}) = J_0(\zeta_0) + \frac{9}{2\pi\sqrt{\lambda m_2}}$ with

$$J_0(\zeta_0) \leqslant \frac{3e}{\sqrt{2(1-\cos\zeta_0)}}\left(1 - \frac{\zeta_0}{\pi}\right)\left(1 \wedge \frac{2}{\lambda m_1 \varrho^*(\zeta_0)}\right), \tag{3.1}$$

and $J_1^{(a)}(\lambda, \boldsymbol{\mu}) = J_1(\zeta_0) + \frac{44}{\pi \lambda m_2}$ with

$$J_1(\zeta_0) \leqslant e\left(1 - \frac{\zeta_0}{\pi}\right)\left\{\left(1 \wedge \frac{2}{\lambda m_1 \varrho^*(\zeta_0)}\right) + \frac{1}{\sqrt{2(1-\cos\zeta_0)}}\left[1 \wedge \frac{4}{\lambda m_1^2 \varrho^*(\zeta_0)}\left(1 + \frac{1}{\lambda \varrho^*(\zeta_0)}\right)\right]\right\}. \tag{3.2}$$

Since $\lambda \asymp n\,p^k\,(1-p+p^k)$, the moments m_1 and m_2 are of order $O(1)$ ($m_1 \asymp 1/(1-p+p^k)$) and the same holds for ζ_0 and $\varrho^*(\zeta_0)$. Thus it remains to analyse the denominators in (3.1) and (3.2). Since $0 < \mu_1 \leqslant 1$, it is easy to see that $\varrho_1^*(\zeta_0) \geqslant (1-p)^2\,(1-\cos\zeta_0) \geqslant (1-p)^2\,(1-\cos(1/(2k-1)^2))$ (using the rough estimate $\zeta_0 > 1/(2k-1)^2$). The same lower bound holds true for $\varrho_2^*(\zeta_0)$ and therefore $\varrho^*(\zeta_0)$ can be bounded from below by $\min(\frac{1}{2}\,(1-p)^2\,(1-\cos(1/(2k-1)^2)), 1) > 0$. Hence $1/\varrho^*(\zeta_0)$ is bounded from above uniformly in n. Moreover, we obtain that $\lambda\, m_2 \geqslant \lambda\, m_1$ and $\lambda\, m_1^2 \geqslant \lambda\, m_1$. Hence the order of $J_1^{(a)}(\lambda, \boldsymbol{\mu})$ depends on the order of $\lambda\, m_1$. A similar discussion yields that the order $J_0^{(a)}(\lambda, \boldsymbol{\mu})$ depends on the order of $\sqrt{\lambda\, m_2}$. Using these bounds for the terms in the denominators of (3.1) and (3.2) as well as the uniform lower bound for ζ_0 we get that the order of $J_0^{(a)}(\lambda, \boldsymbol{\mu})$ is $O(1/(n\,p^k)^{1/2})$ and $J_1^{(a)}(\lambda, \boldsymbol{\mu})$ is of order $O(1/(n\,p^k))$. Since $\varepsilon_1 = O(k\,n\,p^{2k})$, the second summand in (2.6) has order $O(k\,p^k)$. To calculate the order of the remaining contribution (the third summand in (2.6)), we apply Janson's inequality (2.11): with $\Gamma_\alpha^+ = \Gamma_\alpha^{vs}$ we obtain in the k-run example $\delta = 2\sum_{i=1}^{k-1} p^i = O(p)$ and get

$$\mathbb{P}\big(W \leqslant \lambda m_1(1 - \varrho^*(\zeta_0)/4)\big) \leqslant \exp\left(-\lambda m_1(\varrho^*(\zeta_0)/4)^2 \frac{1}{1 + 2\sum_{i=1}^{k-1} p^i}\right).$$

Summarizing, we find that in *Kolmogorov distance* the right-hand side of (2.6) has the order

$$O\big(kp^k + \exp(-\mathrm{const}(k)\, np^k)\big) \tag{3.3}$$

for some $\mathrm{const}(k) > 0$, which is of the good order $O(kp^k)$ as soon as np^k is at all large.

The (less explicit) bound in *total variation* is only available in the case $\lambda \geqslant 2$. The assumption $R(\boldsymbol{\mu}) > 1$ is fulfilled. We only check the influence on order of $C_1(\boldsymbol{\mu})$ since we take $\varepsilon_0 = 0$. To make our explanation more self-contained we state the expressions for $C_1(\boldsymbol{\mu})$ given in [7]: define $m_{[j]} = \sum_{i \geqslant 1} i_{[j]} \mu_i$ with $i_{[j]} = i(i-1)\cdots(i-j+1)$, $\gamma = m_{[2]} + m_1/2$, $m_* = \sup_{j \geqslant 4}\big(\frac{m_{[j]}}{(j-4)!}\big)^{1/j}$, $d_* = (m_*^3 \vee 3m_1)/2$, $\eta_1 = \gamma^{-1}\, m_*$ and $\eta_2 = \gamma^{-1} d_*$. Pick $\zeta_0 = \sqrt{m_2/m_4}$ and $1 < r < R(\boldsymbol{\mu})$ small enough (such that some technical assumptions are fulfilled – see (1.24) and (1.25) in [7]); then we can state the value taken by $C_1(\boldsymbol{\mu})$:

$$C_1(\boldsymbol{\mu}) = \frac{1}{\gamma}\big((130 + 15\eta_1) + \eta_2(232 + 19\eta_1)\big) + \frac{1}{m_1}\left(28 + \frac{12\,e\,r}{(r-1)\varrho^*(\zeta_0)} + 4\sqrt{\frac{\pi}{m_1}}\left(1 + \frac{7.2\,m_3}{m_1^3}\right)\right) + \frac{1}{(r-1)\gamma}\left(4r\sqrt{\frac{m_2}{m_4(r-1)}} + (r-1)^{1/2}\big(-\log(r-1) + 2 + 9\eta_1\big)\right).$$

In the following we briefly discuss the dependence of $C_1(\boldsymbol{\mu})$ through the η_i, $i = 1, 2$. It will come out that in (2.4) we can choose c_1 fixed for all n and that $C_1(\boldsymbol{\mu})$ is bounded uniformly in n. Using quite rough estimates we obtain that $1 \leqslant m_{[j]} \leqslant m_j \leqslant (2k-1)^j$, $3/2 \leqslant \gamma \leqslant (2k-1)^2$, $m_* \leqslant 2k-1$, $d_* \leqslant 4k^3 + O(k)$, $\eta_1 \leqslant (4/3)k$ and $\eta_2 \leqslant (8/3)k^3$. Therefore we obtain the (rough) estimate $C_1(\boldsymbol{\mu}) \leqslant (\text{const}\, k^3)$ with some quite large constant. This calculation shows that we result in a change of the dependence on k in (3.3). Hence $H_0^{(a)}(\lambda, \boldsymbol{\mu})$ is of order $O(c(k)/(n\,p^k)^{1/2})$ with some constant $c(k)$. The leading term $H_1^{(a)}(\lambda, \boldsymbol{\mu})$ is of order $O(k^3/(n\,p^k))$. The main significance is that the constants $C_1(\boldsymbol{\mu})$ exist. They are not chosen optimally. We use the fact that $C_2(\boldsymbol{\mu}) < 1$ to choose c_1 suitably and, therefore, we can apply (2.4) and bound $\mathbb{P}\big(W \leqslant \frac{1}{2}(1 + c_1)\lambda m_1\big)$ again by Janson's inequality and obtain an order of the bound in total variation distance similar to (3.3), possibly with less favourable constants, and under the condition that $\lambda \geqslant 2$. Our estimates for $C_1(\boldsymbol{\mu})$ are extremely rough, which can be seen by discussing the case that p is near 1. There we obtain that m_j is roughly equally like $(2k-1)^j$ and γ is roughly $(2k-1)^2$. Moreover, $C_1(\boldsymbol{\mu})$ is dominated by $1/m_1$, which is roughly $1/(2k-1)$. Apart from constants we conclude that, for p large, the total variation bound could be $O\big(p^k + \exp(-\text{const}\, np^k)\big)$.

Note that, apart from the condition $\lambda \geqslant 2$, there is no other restriction on the value p. We see that this bound is quite applicable when $\mathbb{E}W = np^k \to \infty$, which is not the case for the bounds in [2] and [10].

3.2. Isolated vertices in the rectangular lattice on the torus

Consider a rectangular lattice on the torus with n vertices and $N = 2n$ edges (see [12, Section 3]). Assume that the edges can be deleted independently of each other with a constant probability $1 - p = q$. Label the vertices of the lattice by $\{1, 2, \ldots, n\}$. Define a family $(Q(\alpha), \alpha \in \Gamma)$ of subsets of four edges of the lattice having one common vertex. Index these sets of edges by the label of the central vertex. Note that in the pattern considered the edges, and not vertices, are important. The edges can work or be failed and introduce randomness in the model. Vertices are only used to index the edge-patterns.

Thus $|\Gamma| = n$ and we define $I_\alpha = I[v_\alpha \,\text{is isolated}]$ and $W = \sum_{\alpha\in\Gamma} I_\alpha$. With the above definitions $\mathbb{P}(v_\alpha \,\text{is isolated}) = q^4$ and $\mathbb{E}W = nq^4$. The random variable W counts the number of isolated vertices in the graph. Choose $\Gamma_\alpha^{vs} = \{\beta \neq \alpha : v_\alpha \,\text{and}\, v_\beta \,\text{are neighbours}\}$ and Γ_α^0 and Γ_α^w as in the dissociated partition. By construction and the choice of Γ_α^{vs} the random variables $(I_\alpha, \alpha \in \Gamma)$ are dissociated. Applying Theorem 2.3 we obtain $\varepsilon_1 = 21nq^8$ and $\lambda_i = \frac{n}{i}\binom{4}{i-1} q^{3i+1}(1 - q^3)^{5-i}$, for $i = 1, \ldots, 5$. The second summand in (2.6) has order $O(q^4)$, which we obtain by the same calculations as in the k-run example, and using lower bounds like $\mu_1 = nq^4(1-q^3)^4/\lambda \geqslant (1-q^3)^4 > 0$, $\zeta_0 \geqslant \sqrt{(1 + 4q^3)/(1 + 28q^3 + 72q^6 + 24q^9)}$, as well as the fact that m_4 is finite, $\lambda = O(nq^4)$ ($\lambda = nq^4 - 2nq^7 + 2nq^{10} - nq^{13} + nq^{16}/5$) and the moments m_i are of order $O(1)$. Via Janson's inequality (2.11) with $\delta = 4q^3$ we obtain in Kolmogorov distance that the upper bound is of order

$$O\big(q^4 + \exp(-\text{const}\, nq^4)\big). \tag{3.4}$$

Since $R(\boldsymbol{\mu}) > 1$, the same order holds in total variation distance when $\lambda \geqslant 2$. We omit the calculation of $C_1(\boldsymbol{\mu})$. It follows from [12, Section 3] that an order $O\big((\log^+ 2nq^4)q^4\big)$ of the upper bound for $q^3 \leqslant 1/5$ could be expected when the local approach of [12] would be applied directly. The restrictive condition $q^3 \leqslant 1/5$ is needed there to ensure that $i\lambda_i \searrow 0$. For any q another compound Poisson approximation was suggested by Roos in [12]. Her $CP(\boldsymbol{\lambda}^*)$ can be expressed in our notation as $CP(\lambda^*, \boldsymbol{\mu}^*)$ for a particular choice of parameters suggested in Theorem 2.5 with $l = 2$. In such a case $\sum_{i=3}^{5} i(i-1)\,\lambda_i = 12\,n\,q^{10}\big(1 + O(q^3)\big)$. By (2.9) in Theorem 2.5 we obtain for $d_{TV}\big(\mathscr{L}(W), CP(\lambda^*, \boldsymbol{\mu}^*)\big)$ the same order as in (3.4). Note that our result improves the one of [12] as she was only able to prove that the order of the upper bound on $d_{TV}\big(\mathscr{L}(W), CP(\lambda^*, \boldsymbol{\mu}^*)\big)$ is $O\big((\log^+ 2\,n\,q^4)\,q^4\big)$.

3.3. Two-dimensional consecutive-k-out-of-n system

Assume that a system consists of n^2 components placed on a square grid of size n and it fails if and only if there exist at least m possibly overlapping square sub-grids of size k ($1 < k < n$) with all k^2 components failed. $\Gamma = \{(r,s) : 1 \leqslant r,s \leqslant n-k+1\}$ and for $\alpha \in \Gamma$ denote by $A_\alpha = A_{rs}$ the $k \times k$ sub-grid with left lower-most component (r,s). We do not place it on a torus. Consequently, we have to take boundaries into account. Define a family of subsets $(Q(\alpha), \alpha \in \Gamma)$ of the set $\{1,2,\ldots,n^2\}$ consisting of the points of a $k \times k$ sub-grid and identify each index with the left lower-most component. Define $I_\alpha = I$ (all items in A_α are failed) for each $\alpha \in \Gamma$ and $W = \sum_{\alpha \in \Gamma} I_\alpha$. To define the neighbourhood of very strong dependence we have the choice for the extent of overlap R: for $1 \leqslant R \leqslant k^2$ we can define

$$\Gamma_\alpha^{vs}(R) = \{\beta \in \Gamma \setminus \{\alpha\} : |\beta \cap \alpha| = r, \quad r = R, \ldots, k^2 - 1\}. \tag{3.5}$$

First of all we take $R = k^2 - k$. If the failure probability is q for all items in the grid, we obtain $\mathbb{E}W = (n-k+1)^2 q^{k^2}$ and $|\Gamma_\alpha^{vs}(k^2-k)| \leqslant 4$ and $|\Gamma_\alpha^{vs}(k^2-k)| + |\Gamma_\alpha^{0}(k^2-k)| \leqslant (2k+1)^2 - 1$. Some calculations lead to the following bound for (2.7):

$$\varepsilon_1 \leqslant (n-k+1)^2 q^{k^2} \left((4k^2 + 12k - 3) q^{k^2} + 4 \left(\sum_{u=1}^{k-1} \sum_{v=1}^{k-1} q^{k^2 - uv} + \sum_{v=1}^{k-2} q^{k^2 - kv} \right) \right),$$

which is of order $O\big(q^{2k-1}(n-k+1)^2 q^{k^2}\big)$. Moreover, the λ_is can be computed explicitly (see [10, Section 3.7]): $\lambda_i = \frac{q^{k^2}}{i}\big(4\binom{2}{i-1} q^{k(i-1)}(1-q^k)^{3-i} + 4(n-k-1)\binom{3}{i-1} q^{k(i-1)}(1-q^k)^{4-i} + (n-k-1)^2 \binom{4}{i-1} q^{k(i-1)}(1-q^k)^{5-i}\big)$ for $i = 1, \ldots, 5$. We obtain $\lambda = O\big((n-k+1)^2\, q^{k^2}\big)$. A lower bound is $\lambda \geqslant (n-k-1)^2\, q^{k^2}\,(1 - 2\,q^k + 2\,q^{2k} - q^{3k} + q^{4k}/5)$. The δ in Janson's inequality can be bounded by constant multiplied with q^k and we obtain in total variation distance that the upper bound is of order $O\big(q^{2k-1} + \exp(-\mathrm{const}\,(n-k+1)^2 q^{k^2})\big)$, whenever $(n-k+1)^2\, q^{k^2} \geqslant 2$. The argument presented here provides an improvement on the order $O\big(q^{2k-1} \log^+((n-k+1)^2 q^{k^2})\big)$ obtained in [10, Theorem 3.G]. One might be interested in taking $R = 1$ (maximal clump) instead of $R = k^2 - k$. It would imply ε_1 to be of order $O\big((n-k+1)^2\, q^{2k^2}\big)$. Thus the order of the upper bound would be $O\big(q^{k^2} + \exp(-\mathrm{const}\,(n-k+1)^2 q^{k^2})\big)$. However, in such a case the computation of the parameters λ_i would be very laborious.

3.4. U-statistics

Suppose that Γ is a collection of k-subsets $\alpha = \{\alpha_1, \ldots, \alpha_k\}$ of $\{1,2,\ldots,n\}$ and define $X_\alpha = \phi_\alpha(Y_{\alpha_1}, \ldots, Y_{\alpha_k})$ for some nonnegative integer-valued symmetric (in its arguments) functions ϕ_α, where $Y_1, \ldots, Y_n$ are independent random elements of some space $\mathscr{X}$. Then $W = \sum_{\alpha \in \Gamma} X_\alpha$ is called an *incomplete* U-statistic. If Γ is the set of all k-subsets, W is called a U-statistic. Many statistics commonly used are in fact members of this class (see [9]). Given $R \geqslant 1$, similarly to (3.5), the neighbourhoods of dependence can be defined by $\Gamma_\alpha^{vs}(R) = \{\beta : R \leqslant |\beta \cap \alpha| \leqslant k-1\}$. Thus, $\Gamma_\alpha^0 = \{\gamma : |\gamma \cap \beta| \geqslant 1$ for some $\beta \in \{\alpha\} \cup \Gamma_\alpha^{vs}(R)\} \setminus \{\{\alpha\} \cup \Gamma_\alpha^{vs}(R)\}$. The choice of R affects the order of the term $\mathbb{E}(X_\alpha U_\alpha)$, which is frequently that of the largest order. Let us calculate ε_1 for W depending on R. Denote $\mathbb{E}(X_\alpha) = \pi_\alpha$. For $1 < R < k$ the terms of (2.7) are of the following form:

$$\sum_{\beta \in \Gamma_\alpha^0(R)} \mathbb{E}(X_\alpha X_\beta) = \sum_{r=1}^{R-1} \sum_{\substack{\beta \in \Gamma_\alpha^0(R) \\ |\beta \cap \alpha| = r}} \mathbb{E}(X_\alpha X_\beta) + \sum_{\substack{\beta \in \Gamma_\alpha^0(R), |\beta \cap \alpha| = 0 \\ |\gamma \cap \beta| \neq 0, \gamma \in \Gamma_\alpha^{vs}(R)}} \pi_\alpha \pi_\beta,$$

and

$$\sum_{\beta \in \Gamma_\alpha^{vs}(R) \cup \Gamma_\alpha^0(R)} \pi_\alpha \pi_\beta = \sum_{r=1}^{k-1} \left(\sum_{\substack{\beta \in \Gamma_\alpha^{vs}(R) \cup \Gamma_\alpha^0(R) \\ |\beta \cap \alpha| = r}} \pi_\alpha \pi_\beta \right) + \sum_{\substack{\beta \in \Gamma_\alpha^0(R), |\beta \cap \alpha| = 0 \\ |\gamma \cap \beta| \neq 0, \gamma \in \Gamma_\alpha^{vs}(R)}} \pi_\alpha \pi_\beta.$$

Consider the case $\pi_\alpha = \pi$ for all $\alpha \in \Gamma$. Let $n_r(\alpha) = |\{\beta : |\beta \cap \alpha| = r\}|$, $r = 1, \ldots, k-1$, $m(\alpha, R) = |\{\beta \in \Gamma_\alpha^0(R) : |\beta \cap \alpha| = 0, |\gamma \cap \beta| \neq 0$ for some $\gamma \in \Gamma_\alpha^{vs}(R)\}|$ and $\sigma_r = \mathbb{E}(X_\alpha X_\beta)$ for all pairs α, β satisfying $|\alpha \cap \beta| = r$. Then we obtain

$$\varepsilon_1 = \sum_{\alpha \in \Gamma} \left(\pi^2 + \sum_{r=1}^{R-1} n_r(\alpha) \sigma_r + m(\alpha, R) \pi^2 + \sum_{r=1}^{k-1} n_r(\alpha)\, \pi^2 + m(\alpha, R) \pi^2 \right).$$

Let us consider the case $|\Gamma| = \binom{n}{k}$. We obtain $|\{\beta : |\beta \cap \alpha| = r\}| = \binom{k}{r}\binom{n-k}{k-r}$, which is of order $O(n^{k-r})$. The size of the strong dependence region is of order $O(n^{k-R})$. Moreover, $|\Gamma_\alpha^0| = \binom{n}{k} - \sum_{r=R}^{k-1} \binom{k}{r}\binom{n-k}{k-r}$, so $\Gamma_\alpha^w = \emptyset$. For simplicity we assume that the constants $C_l(\lambda, \boldsymbol{\mu})$ in the total variation distance bound do not influence the order of the bound (as we have observed in all our other examples) and that $\boldsymbol{\mu}$ satisfies the assumption $R(\boldsymbol{\mu}) > 1$, $\lambda \geqslant 2$ and fulfils Condition 2.2. Moreover, apply Chebyshev's inequality to establish that

$$\mathbb{P}\big(W \leqslant 1/2(1+c_1)\lambda m_1\big) \leqslant \lambda\, m_2 \big(1/2(1-c_1)\lambda\, m_1\big)^{-2} = O(1/\lambda)$$

(use Bernstein-type or Hoeffding-type inequalities for U-statistics for more refinement). Therefore, for complete U-statistics we find that

$$d_{TV}\big(\mathscr{L}(W), CP(\lambda, \boldsymbol{\mu})\big) \leqslant \mathrm{const}(\boldsymbol{\mu})\, \pi \binom{n}{k} + \sum_{r=1}^{R-1} \frac{\sigma_r}{\pi} n_r(\alpha) = O\left(\pi n^k + \sum_{r=1}^{R-1} \frac{\sigma_r}{\pi} n^{k-r} \right).$$

The order for Poisson approximation is $O\big(\pi n^{k-1} + \sum_{r=1}^{k-1} \frac{\sigma_r}{\pi} n^{k-r}\big)$ (see [6, Section 2.3, (3.2)]). This result is not an improvement in general. Consider the following example: $k = 2$ and $\phi(x, y) = \phi_\alpha(x, y) = x\,y$ and Y_1 is an indicator with $\mathbb{P}(Y_1 = 1) = p$. Then the order of Poisson approximation is $p^2\, n + p\, n$, whereas the order of compound Poisson approximation is $p\,n$. In the incomplete case our result is applicable, provided the quantities $n_r(\alpha)$ and $m(\alpha, R)$ are uniformly bounded as Γ grows. The smaller R is, the better is the order of approximation. Examples 3.1 to 3.3 above are incomplete U-statistics. The choice of Γ, the collection of k-subsets, represents the local character shared by all our examples. These observations lead to a natural question: 'how incomplete' should a U-statistics be to get better results when approximating by a compound Poisson distribution rather than a Poisson one.

3.5. Other examples

Note that further examples with a dissociated structure are studied in [6, Chapter 7], [4] and [10]. They deal with connected s-systems and consecutive 2-systems, where coloured graphs and birthday problems come into play. In all these examples our results lead to improved bounds for an appropriately chosen compound Poisson approximation.

References

[1] Aldous, D. J. (1989) *Probability Approximations via the Poisson Clumping Heuristic*, Springer, New York.

[2] Arratia, R., Goldstein, L. and Gordon, L. (1990) Poisson approximation and the Chen–Stein method. *Statist. Science* **5** 403–434.

[3] Barbour, A. D., Chen, L. H. and Loh, W.-L. (1992) Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20** 1843–1866.

[4] Barbour, A. D., Chryssaphinou, O. and Roos, M. (1996) Compound Poisson approximation in systems reliability. *Naval Research Logistics* **43** 251–264.

[5] Barbour A. D. and Eagleson, G. K. (1984) Poisson convergence for dissociated statistics. *J. Royal Statist. Soc.* **B 46** 397–402.

[6] Barbour, A. D., Holst, L. and Janson, S. (1992) *Poisson Approximation*, Oxford University Press, Oxford.

[7] Barbour, A. D. and Utev, S. (1999) Compound Poisson approximation in total variation. *Stoch. Process. Appl.* **82** 89–125.

[8] Barbour, A. D. and Utev, S. (1998) Solving the Stein equation in compound Poisson approximation. *Adv. Appl. Probab.* **30** 449–475.

[9] Lee, A. J. (1990) *U-Statistics: Theory and Practice*, Marcel Dekker, New York.

[10] Roos, M. (1993) *Stein–Chen method for compound Poisson approximation.* PhD thesis, University of Zürich, Switzerland.

[11] Roos, M. (1994) Stein–Chen method for compound Poisson approximation: the coupling approach. *Probab. Theory Math. Statist.* **4** 645–660.

[12] Roos, M. (1994) Stein's method for compound Poisson approximation: the local approach. *Ann. Appl. Probab.* **4** 1177–1187.